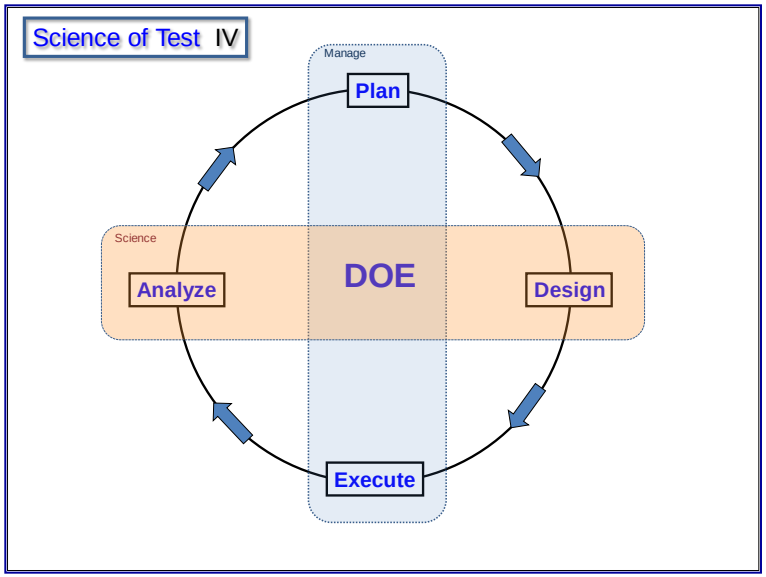


Embedding DOE in Military Testing

One Organization's Roadmap

presented to:

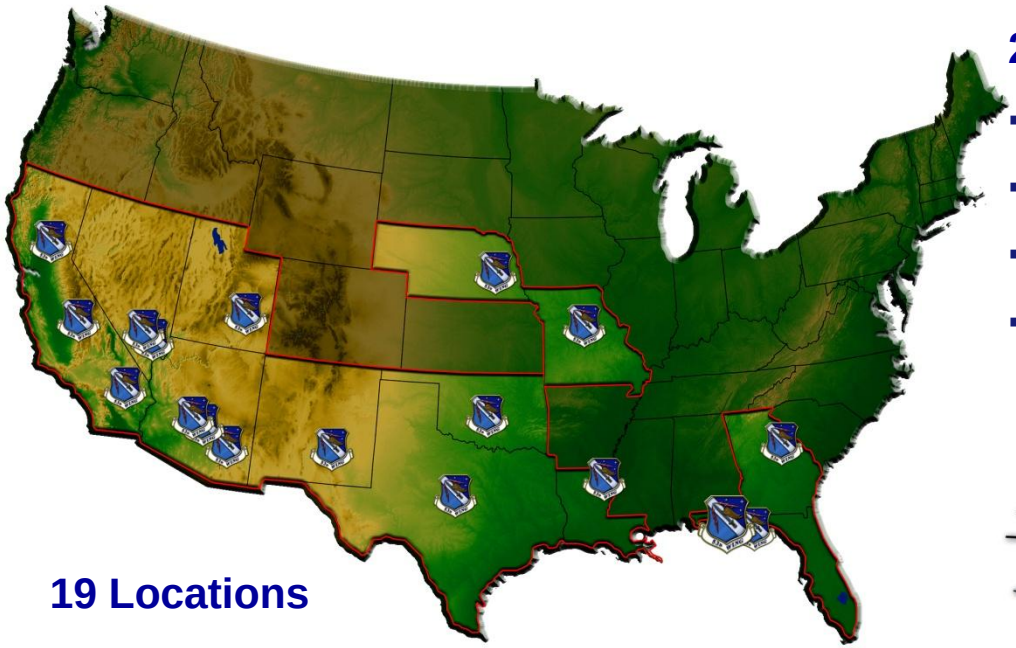
2011 NASA Statistical Engineering Symposium

May 2011

Jim Simpson, 53d Wing – Greg Hutto 46 Test Wing – Alex Sewell 53d Wing

53d Wing

Mission: Develop, test and evaluate advanced weapons, systems and tactics to perfect the lethality, survivability, and sustainability of our nation's combat forces



19 Locations

2000 Professionals consisting of…

- **550 Officers**
- **650 Enlisted**
- **450 Civilians**
- **350 Contractor**

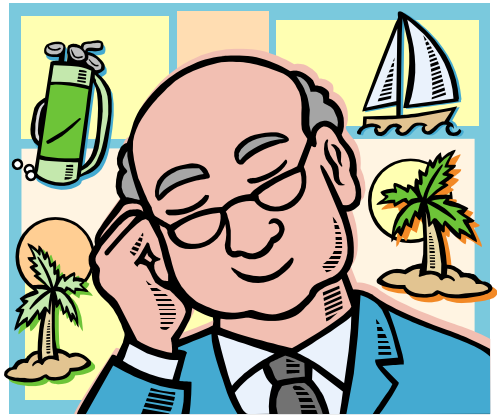
53d Wing
Analysts and Test Engineers

53d Wing

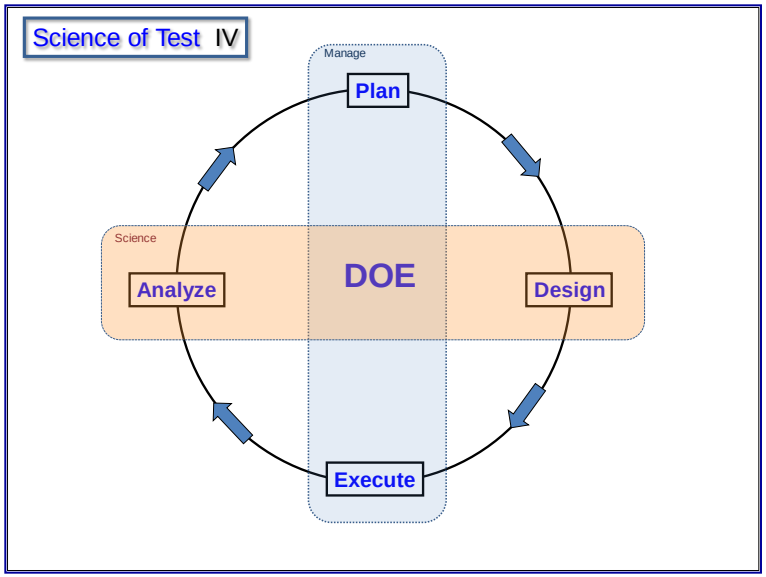
	Electronic Warfare	Weapons Evaluation	Test & Evaluation	Test Management
Analysts 71	7	9	7	48
Test Engineers 124	50	16	20	38

What is *Your* Dream?

"***Be Careful What you Ask for . . .***" **Kevin Burns, Ops Test, Tech Advisor**



DOE is Policy
In DoD



Changing a Culture

Contrast Traditional Methods …

Cases

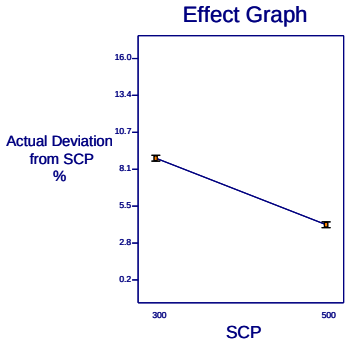
Case	Configuration	Outcome
1		Good
2		Good
3		OK
4		Good
5		Good

Good to go!

OR

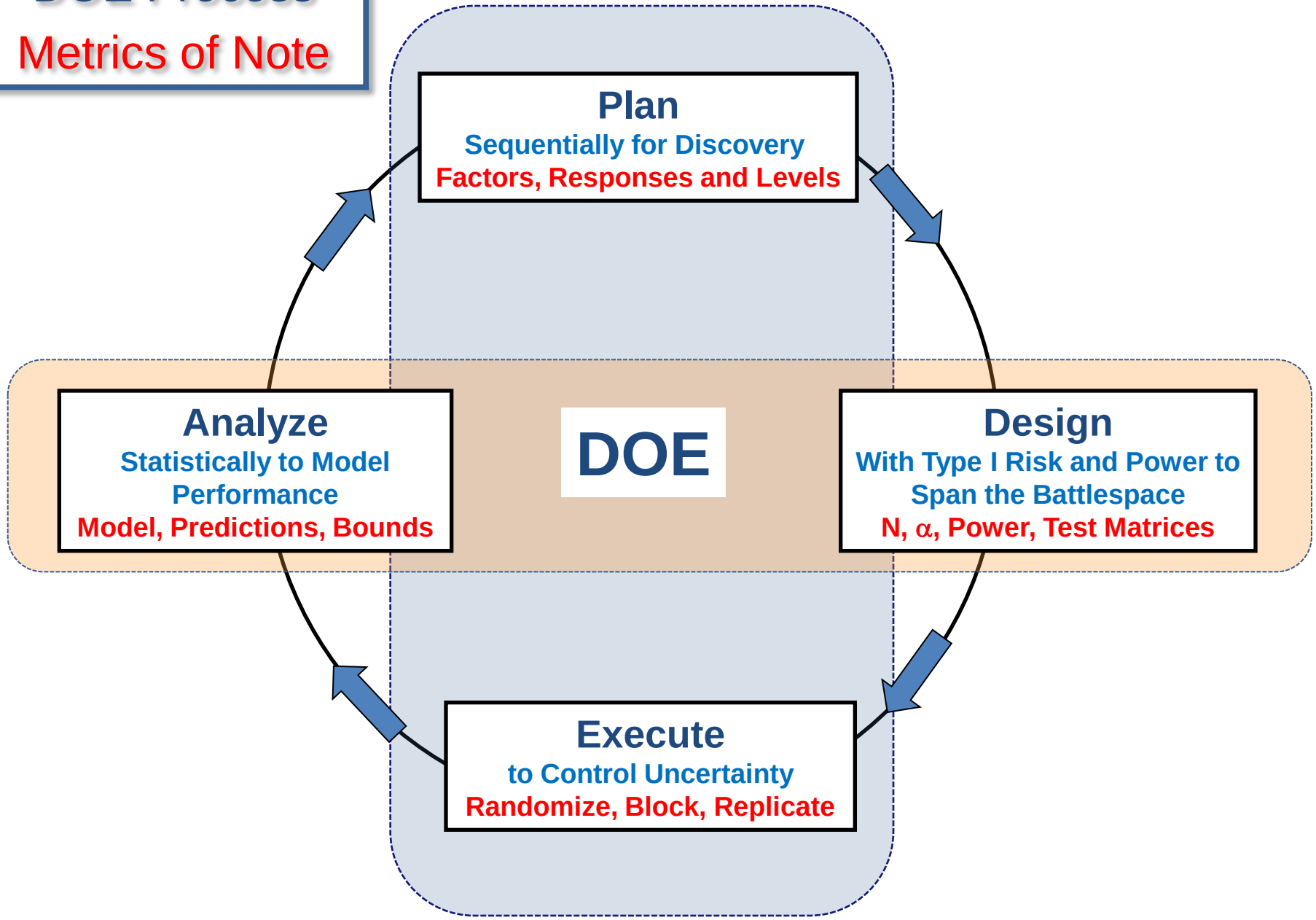
One Factor-at-a-Time

Case	A	B	C
1	**1**	0	0
2	**2**	0	0
3	**3**	0	0
4	**4**	0	0
5	0	**1**	0
6	0	**2**	0
7	0	**3**	0
8	0	**4**	0
9	0	0	**1**
10	0	0	**2**
11	0	0	**3**
…	…	…	…



DOE Process

Metrics of Note

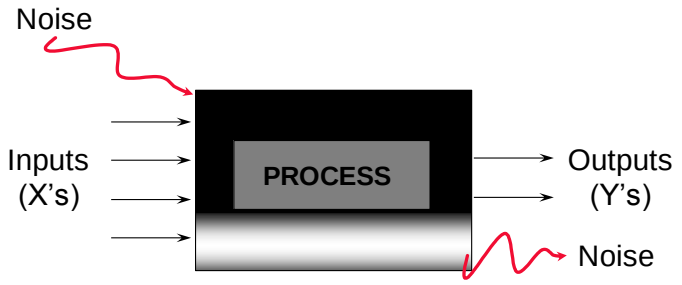


Questions in Testing

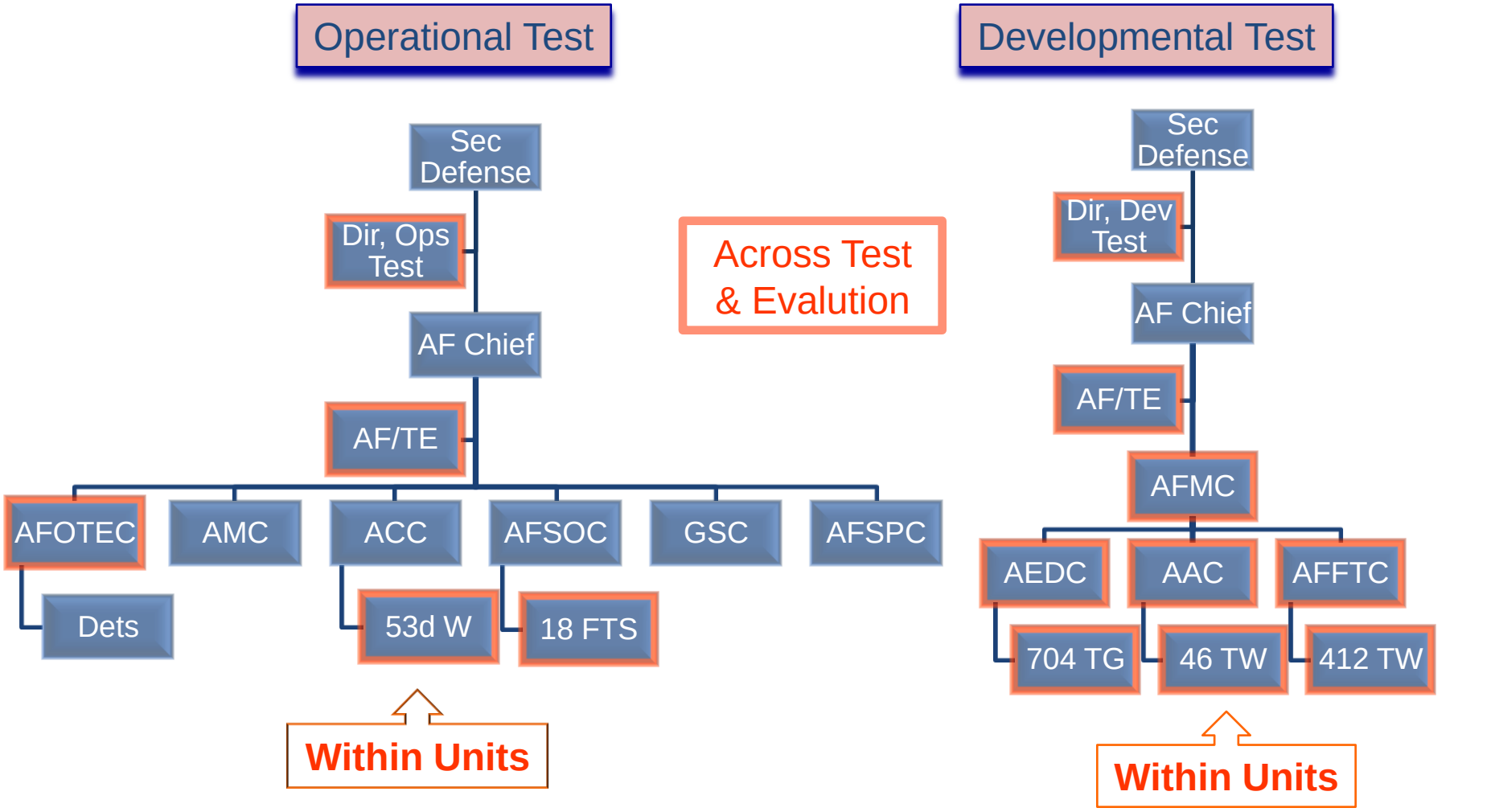
Four Challenges faced by any test

1. *How Many?* A: Sufficient samples to control our twin errors – false positives & negatives
2. *Which Points and What's Good?* A: Span the battle-space with orthogonal run matrices using continuous measures tied to the test objectives
3. *How Execute?* A: Randomize and block runs to exclude effects of the lurking, uncontrollable nuisance variation
4. *What Conclusions?* A: Build math-models of input/output relations, quantifying noise, controlling error

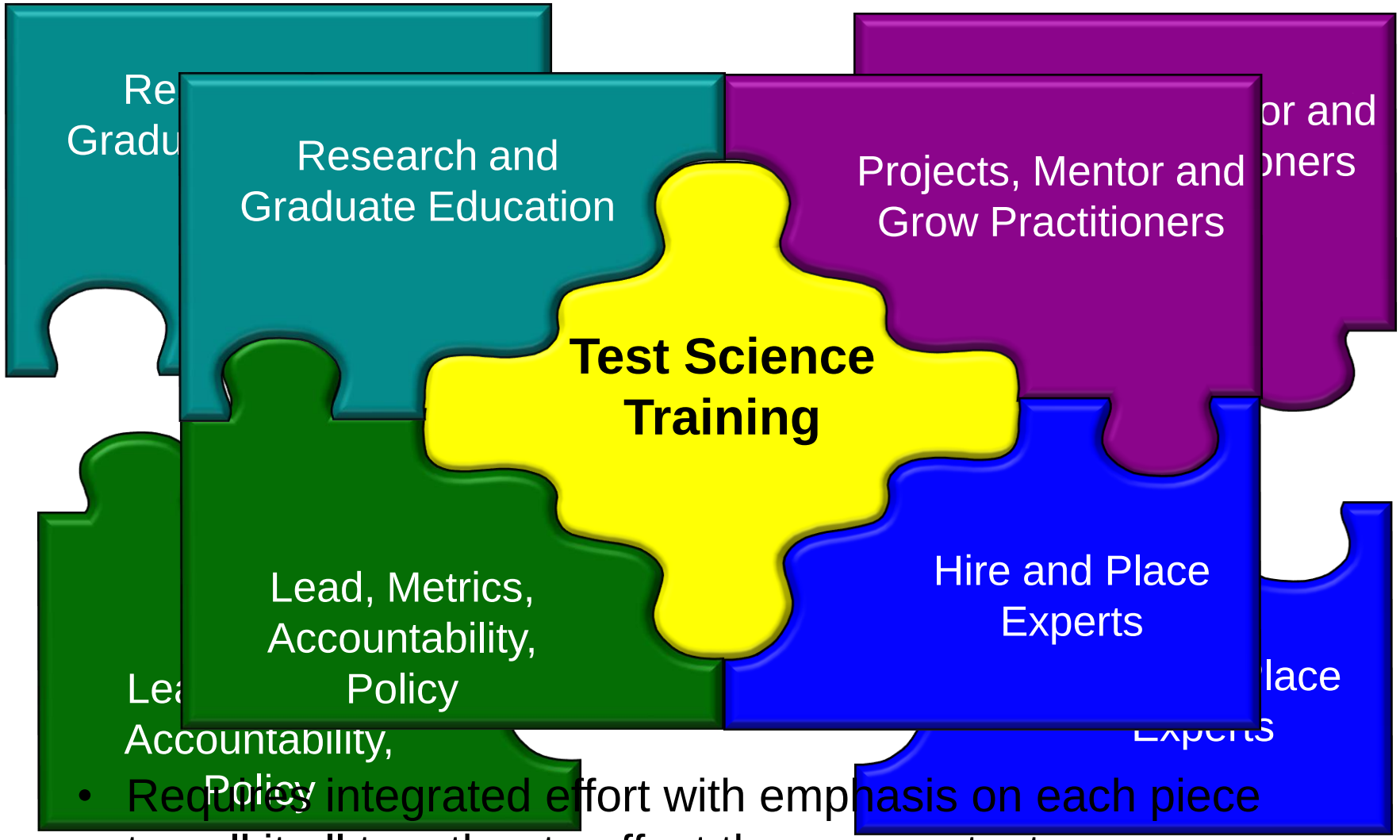
Design of Experiments effectively addresses all these challenges!



Culture Change Across Units



Organization Change Pieces
Move into Place Simultaneously



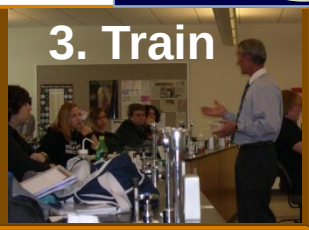
- Requires integrated effort with emphasis on each piece to pull it all together to affect the way we test

Science of Test

Steps to Implementation within Unit

5. Standards

3. Train



4. Mentor

2. Short-Term Wins

I. Leadership --Why DOE?

II. Technical Continuity

1. Foundations

III. Process Improvement

IV. Change Org Structures

Leading the Science of Test

- Stay tuned for the next talk …

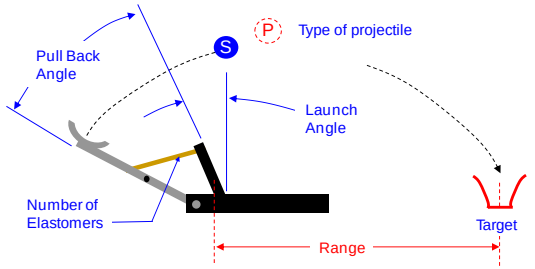
I. Leadership --Why DOE?
II. Technical Continuity

1. Foundations

III. Process Improvement
IV. Change Org Structures

Training our Total Test Team

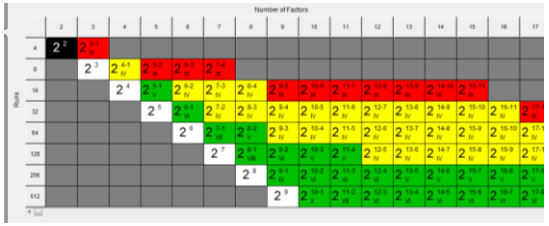
- Leadership, Support and Operator Series
  - DOE Executive Interview (1-2 hour)
  - DOE for Leaders, Aircrew (half day)
  - Intro to Design of Experiments (2 days)
  - DOE Foundations (1 week)



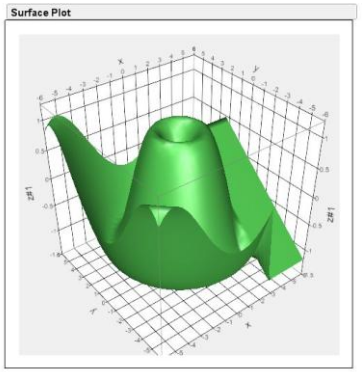
- Analyst and Test Engineer Practitioner Series
  - Each 1-week course uses Discussion-Seatwork-Projects
  - DOE 0 – DOE Foundations for Science of Test
  - DOE I – Design and Analysis of Factorial and Fractionated Designs
  - DOE II – Response Surface Methods, Optimal Designs, Split Plots, Analysis of Ugly Data

Software for Practitioners

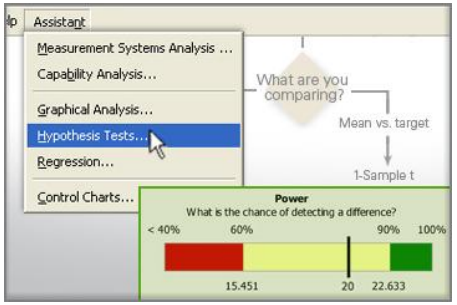
- *Design Expert* – software solely for design of experiments
  - Keeps the analyst focused on DOE procedure
  - Warns when going wayward
  - Used in DOE 0, I, II and in-part III



- *JMP* – general purpose statistical software
  - Industry leader, affordable, requires learning curve
  - Best for our advanced users and needs
  - For DOE III and difficult problems



- *Minitab* – general purpose
  - Interface similar to Excel, user friendly
  - DOE emphasis
  - Split-plot capable



Growing & Mentoring Practitioners

Practitioner* -- (prak-tish-un-ur) n. 1. One who practices an occupation, method or technique.

28 TES/EAA Initial OA Qual Training (ch8 14 Aug 09)

Trainee:	Date complete:				
Training Items	**Method**	**Training Information**	**Date Complete**	**Trainee Initials**	**Trainer Initials**
Wing Directed (mandatory) Training - (complete prior to initial qualification)					
Test Team Training (TTT)	FC	53TMG/TR Webpage			
Design of Experiment (DOE) 0	FC	53TMG/TR Webpage			
DOE I	FC	53TMG/TR Webpage			
DOE II	FC	53TMG/TR Webpage			
53WG 99-103 Review (complete prior to initial qualification)					
Read 53 WGI 99-103 *Capability Based T&E*	SS	53WG Handbook			
Test Team Handbook Review (complete prior to initial qualification)					
Review 53WG Test Team Handbook	SS	53WG Webpage			
Test process checklists	SS	53WG Handbook			
Test template review	SS	53WG Handbook			
Test regulation review	SS	53WG Handbook			
Test Management Review (complete prior to initial qualification)					
TMS Use / Procedures	OJT	PM			
Attend MRR	SS	TBD			
Attend CoT/CRR/FRR	SS	Thursdays (1300)			
Review Test Priority List	SS	TMS			

28 TES/EAA Experienced OA Qual Training (Ch8 14 Aug 09)

Trainee:	Date complete:				
Training Items	**Method**	**Training Information**	**Date Complete**	**Trainee Initials**	**Trainer Initials**
Wing Training - Experienced OA Qual (complete prior to experienced OA qualification)					
Project Management Training (PMT)	FC	TMG Webpage			
Operations Suitability Training (OST)	FC	TMG Webpage			
D0E III	FC	TMG Webpage			
Supplemental Certification Training (complete courses prior to experienced OA qualification)					
ACQ 101 - Fundamentals of Systems Acquisition Management	OL	DAU website			
SYS 101 - Systems Planning, Research, Development & Engineering	OL	DAU website			
Level I T&E Certification					
TST 102 - Fundamentals of Test & Evaluation	OL	DAU website			
CLE 023 - Modeling and Simulation for Test and Evaluation	OL	DAU website			
Level I Program Management Certification					
CLB 007 - Cost Analysis	OL	DAU website			
CLB 016 - Introduction to Earned Value					

- **Various practitioner levels – requires experience**
  - **OA – Initial Qual, Experienced, Instructor**
  - **TE – Initial Qual, Experienced**
- **Include re-qualification**

Long Term Solution
Leadership: Making Changes Endure

DOT&E, DDT&E and Service TE Policy Providing Leadership

MEMORANDUM OF AGREEMENT

SUBJECT: Using Design of Experiments for Operational Test and Evaluation

Regarding the subject, we endorse the enclosed findings of the Operational Test Agency Technical Directors and the Science Advisor for Operational Test and Evaluation.

Dr. Charles E. McQueary
Director, Operational Test & Evaluation

Stephen T. Sargeant, Major General, USAF
Commander, AFOTEC

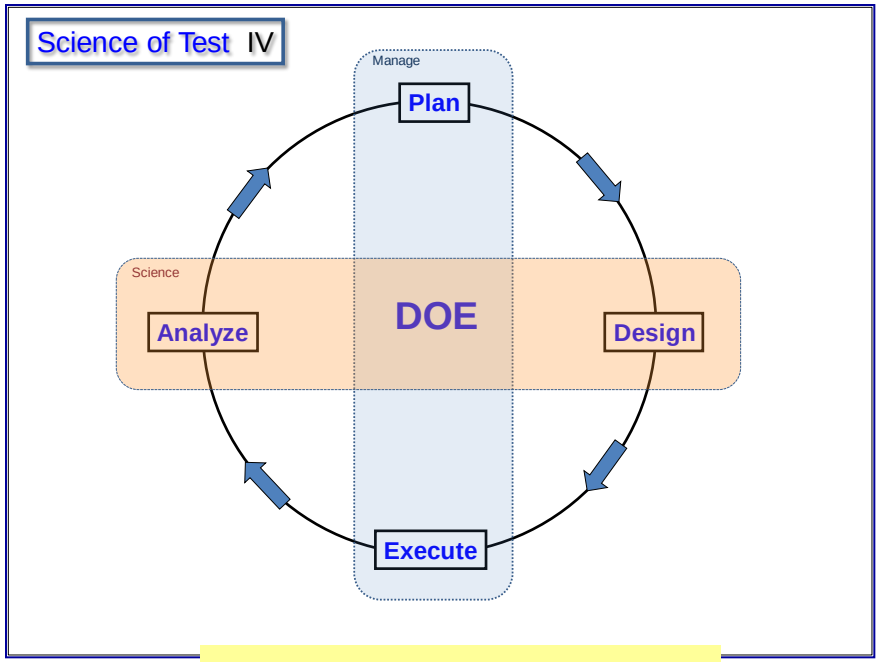
Roger A. Nadeau, Major General, USA
Commander, ATEC

David A. Dunaway, Rear Admiral, USN
Commander, OPTEVFOR

David L. Reeves, Colonel, USMC
Director, MCOTEA

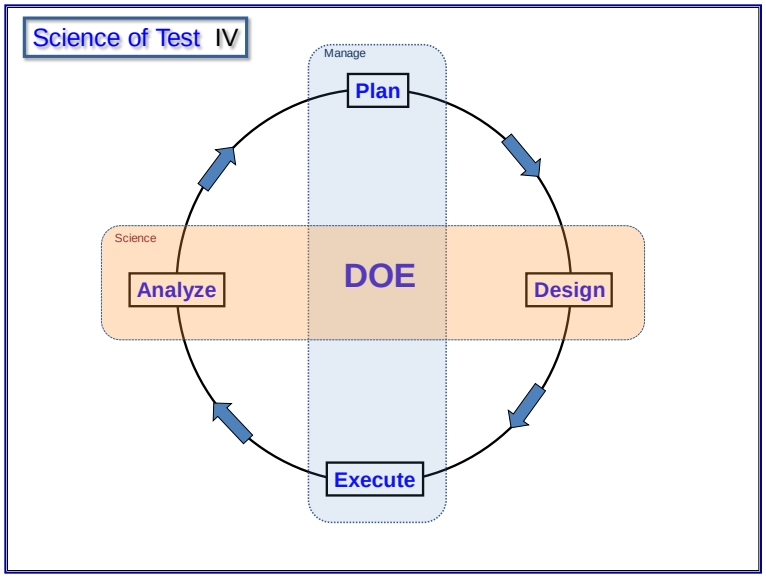
Ronald C. Stephens, Colonel, USA
Commander, JITC

Policy/Guidance



Accountability/Metrics

5. Standards



Defining What We Do

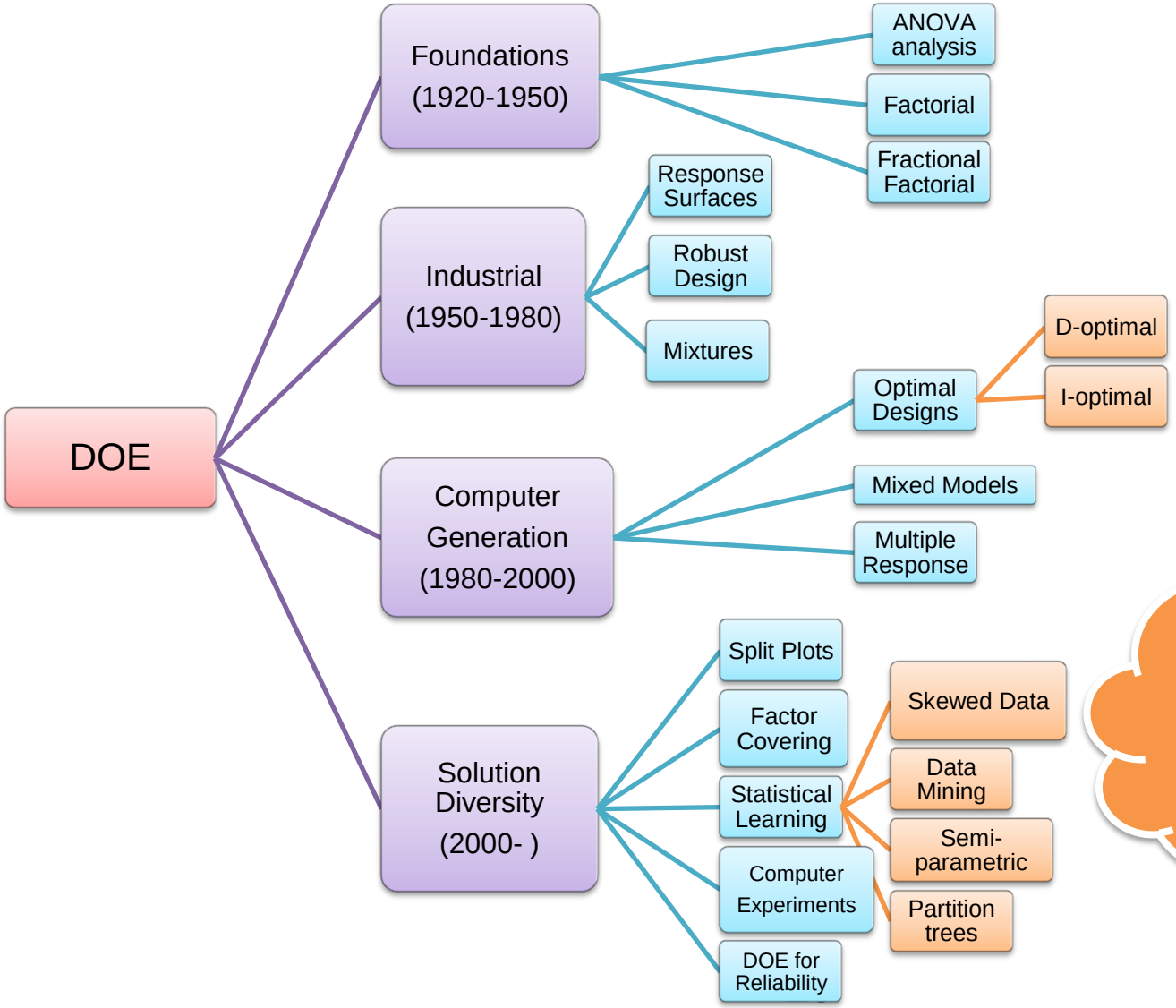
What's in a Name?

- DOE or even Design of Experiments has downside
  - DOE – Energy, Education …?
  - We already design experiments
  - We test, we don't experiment
  - It isn't just DOE, we need a supporting cast of methods
- Label alternatives
  - Operations Analysis, Industrial Statistics
  - Statistical and Probabilistic DOE
  - Statistically Defensible Test
  - Scientific Test and Evaluation Design
  - Test Science or Science of Test
  - Statistical Engineering or Quality Engineering

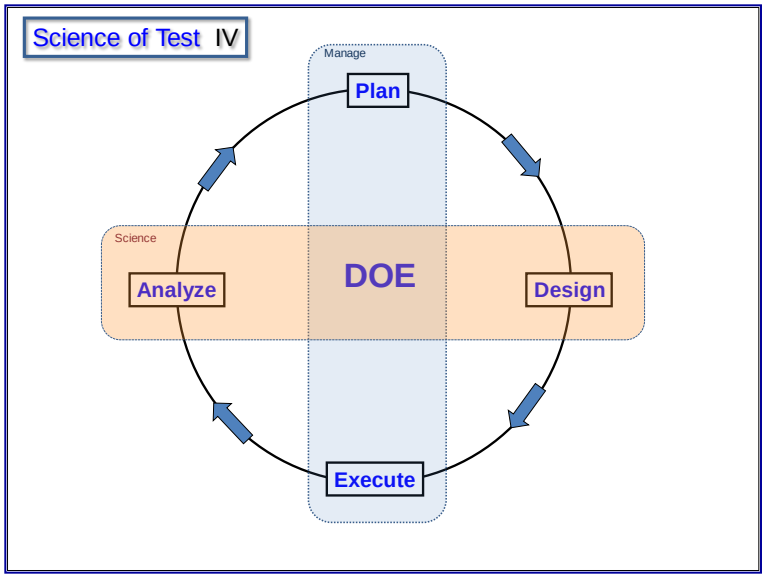
One Term for All Test Science

- DOE is used for planning, design, execution and analysis
- DOE uses statistical, probabilistic, and mathematical (including operations research) methods
- DOE encompasses the entire history of design and statistical techniques and methods peer reviewed and demonstrated effective
- DOE is relevant to all types of testing: developmental and operational, deterministic and high-noise systems, for all system complexities
- DOE is *not* the solution for one-shot proof of concept or demonstrations

DOE Evolution

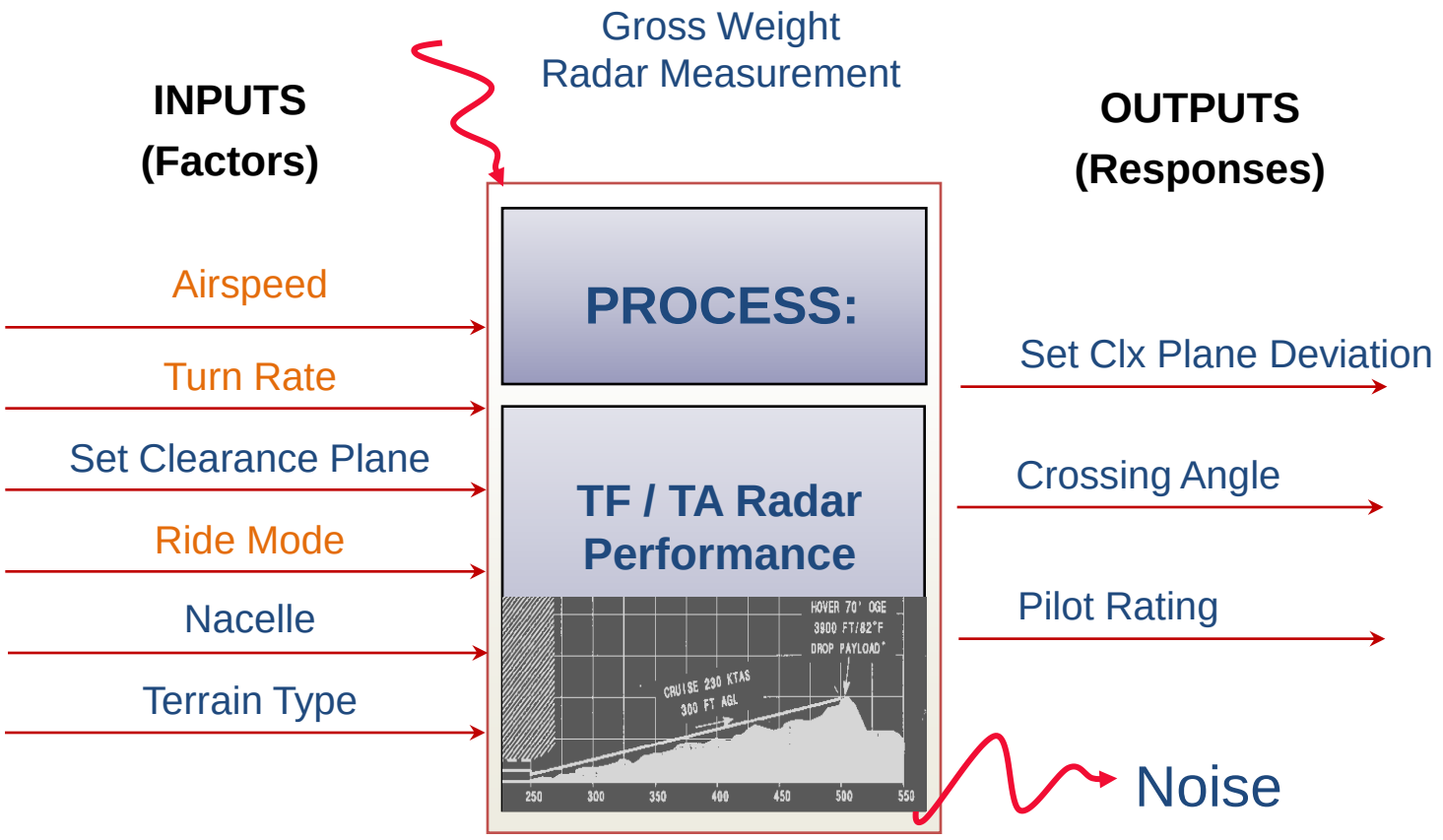


Many of these cross-pollinated from other disciplines



Necessary Tools and Concepts

CV-22 TF Flight Test



$$\text{Responses} = f\left(\text{Factors}\right) + \varepsilon$$

Consider the possible effects of three variables: Airspeed, Turn Rate, and Ride

Risks (α and β) Reviewed

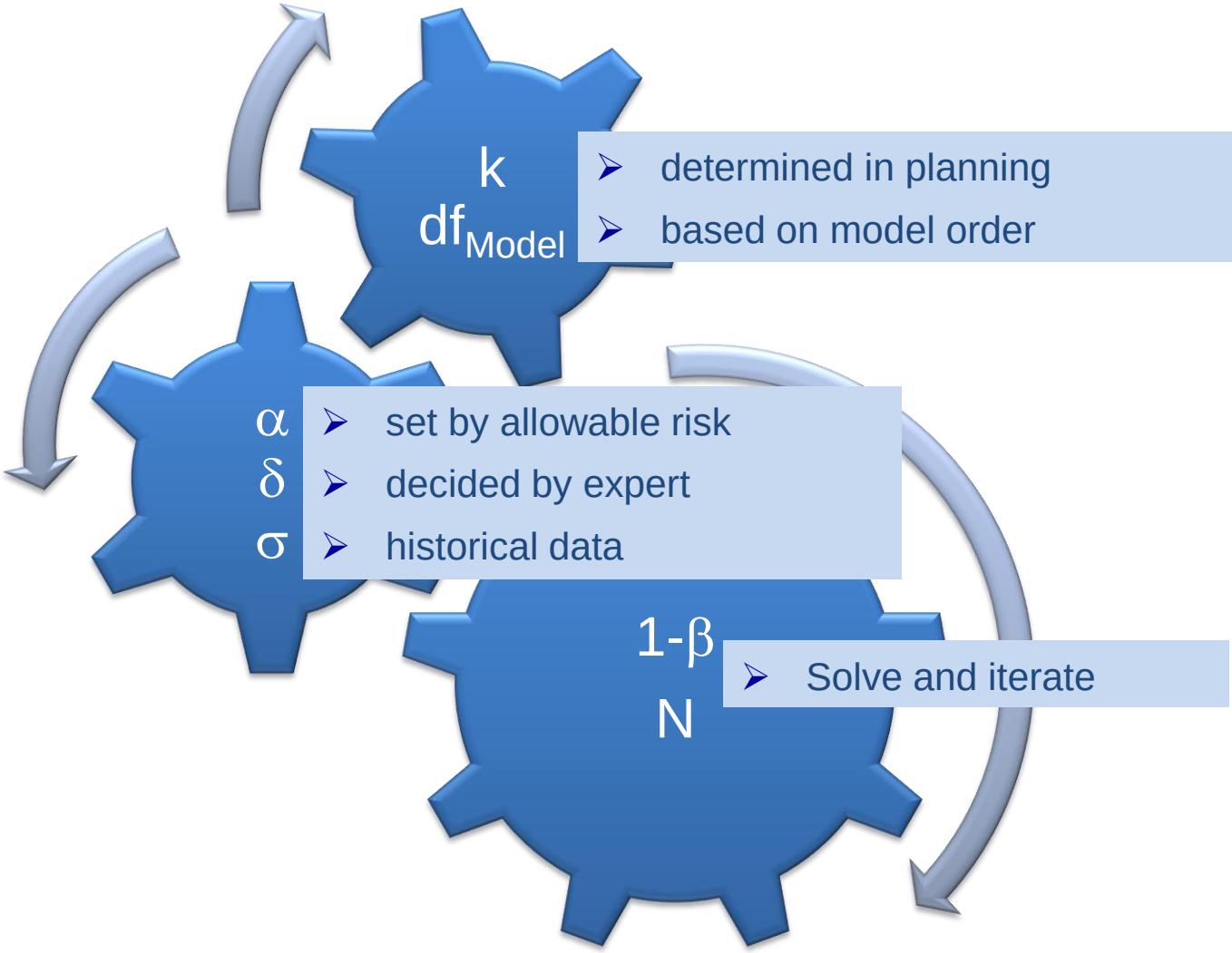
Truth Model: **Response = Ride + Turn**

Test Factors	Hypotheses	Possible Conclusion	Error
A: Airspeed	**H_0: Airspeed has no effect** H_1: Airspeed matters	Airspeed matters	α
B: Turn	H_0: Turn has no effect **H_1: Turn matters**	**Turn matters**	**None,** $1-\beta$
C: Ride	H_0: Ride has no effect **H_1: Ride matters**	Ride has no effect	β

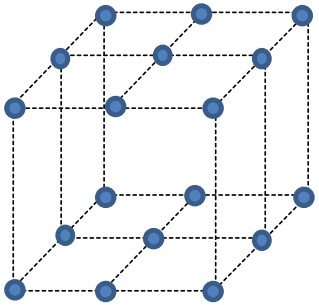
*** Bold Blue reflects the truth**

23

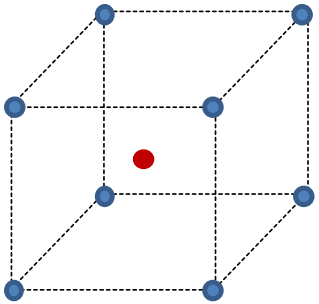
Power Analysis Sequence



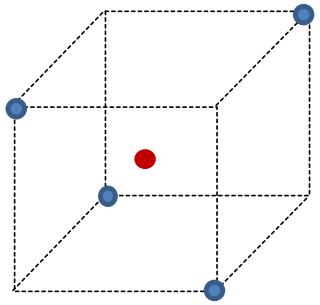
Classic Experimental Designs



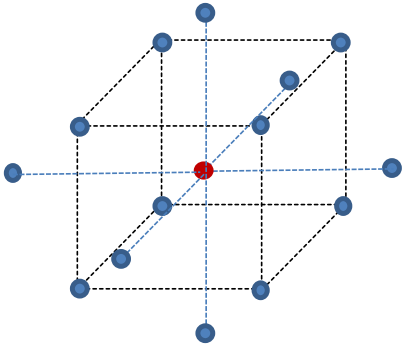
General Factorial
3x3x2 design



2-level Factorial
2^3 design

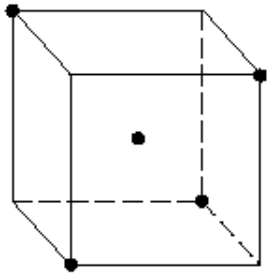


Fractional Factorial
2^{3-1} design

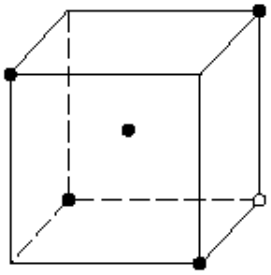


Response Surface
Central Composite design

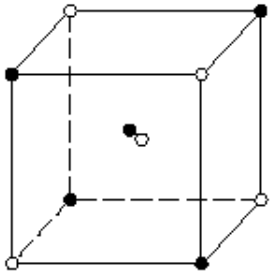
Possible Strategies for Follow-Up Experimentation Following a Fractional Factorial Design



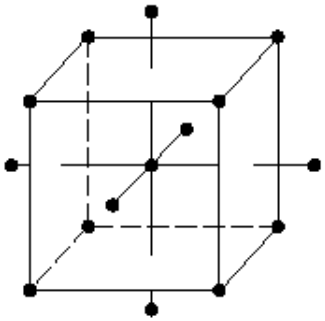
(g) Move to new location to explore an apparent trend in response



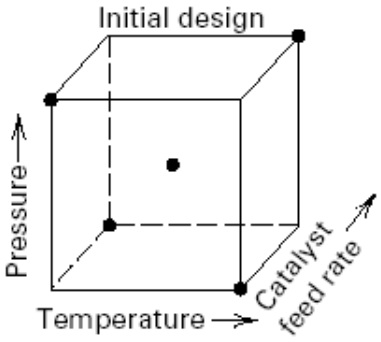
(a) Perform one or more confirmation runs to verify the conclusion from the original fraction

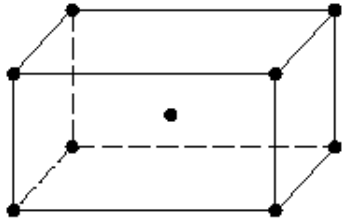


(b) Add another fraction to resolve ambiguities from the original fraction

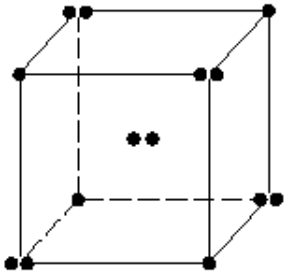


(f) Augment to model apparent curvature

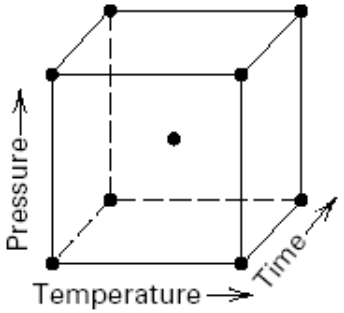




(c) Rescale some factors because they may have been varied over inappropriate ranges



(e) Replicate to improve estimates of effects or because some runs were incorrectly made

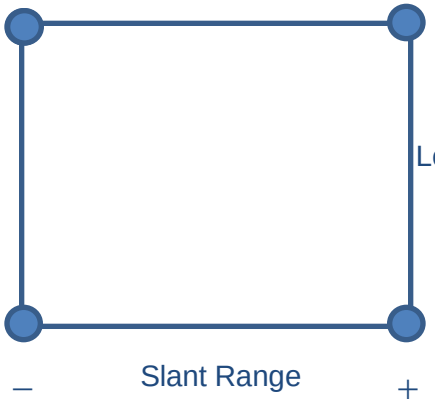


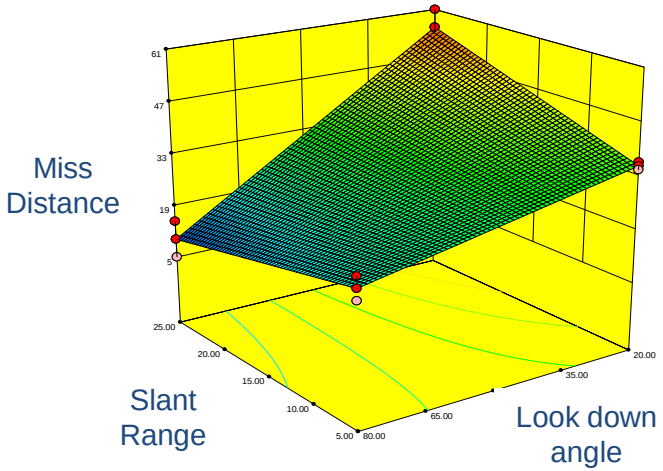
(d) Drop and add factors because the original factor catalyst feed rate is negligible

Adapted from Box, GEP (1992-1993), "Sequential Experimentation and Sequential Assembly of Designs," *Quality Engineering*, Vol 5., No. 2, pp., 321-330.

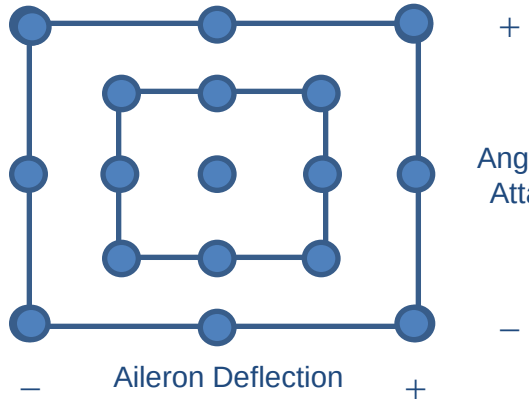
Designs Support the Model

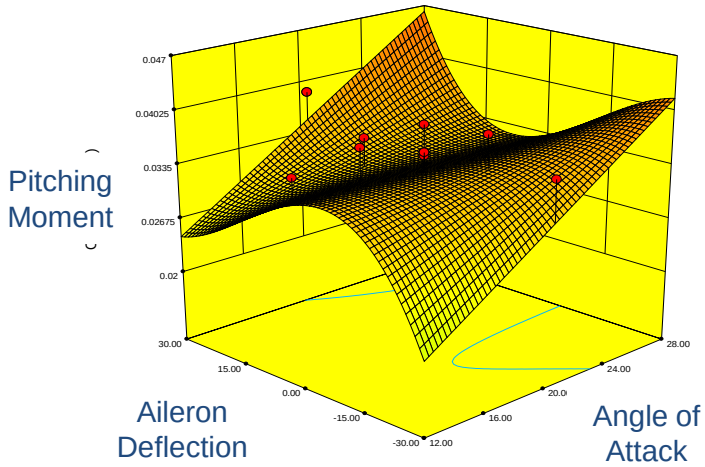
a)





b)





Standard Modeling
Least Squares Regression

Linear in parameters

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon$$

Quantitative
Continuous

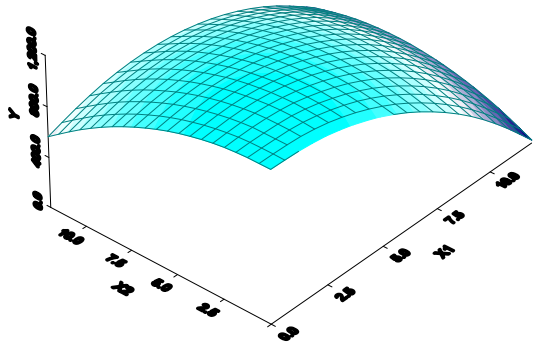
Quantitative
Continuous

Normally distributed
Independent
Homogeneous variance
Single error component

Low correlation

Run	A	B	C
1			
2			
3			
4			
5			

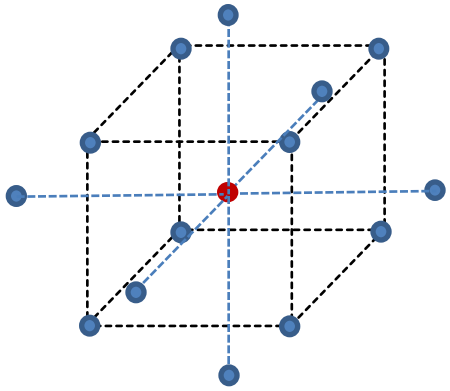
Void of
outliers, leverage points



2nd Order Designs

Assumptions
Randomized
Numeric or Categorical
Mostly Numeric
> 2 level

Design



Attributes
Replication
2nd order design
Nearly Orthogonal
Target Prediction and Coefficient Variance
Efficient runs for k < 7

Assumptions
Errors NID (0, σ^2)
Model is adequate
Y well behaved

Model

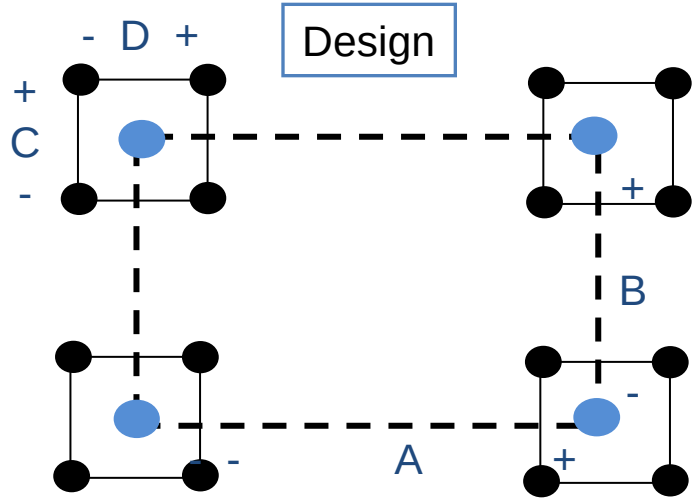
$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i<j} \beta_{ij} x_i x_j + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \varepsilon$$

Attributes
All effects for general model
Pure error + LOF
Nearly Independent β estimates

Split-Plot Designs

Assumptions
Hard to Change Factors
Numeric or Categorical

Design



Attributes
Replication
Orthogonal

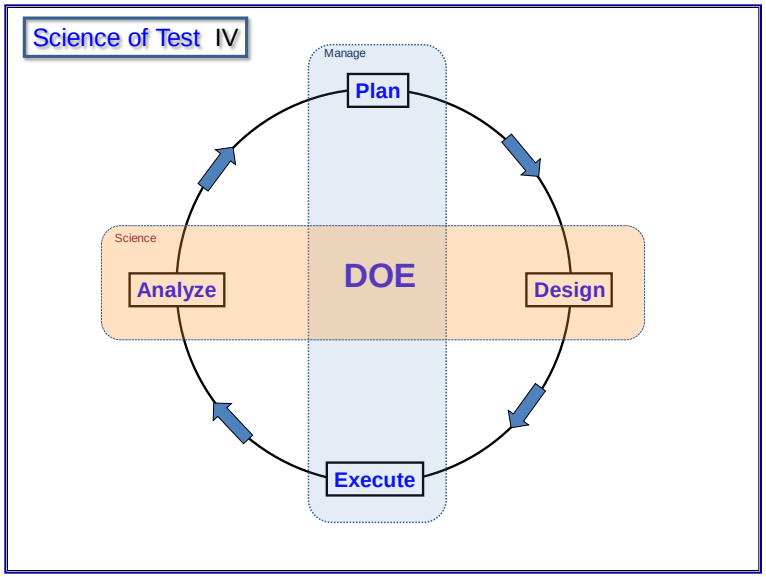
Assumptions
Two Independent Error Terms, both NID (0, σ^2)
Model is adequate
Y well behaved

Model

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i<j} \beta_{ij} x_i x_j + \delta + \varepsilon$$

WP error

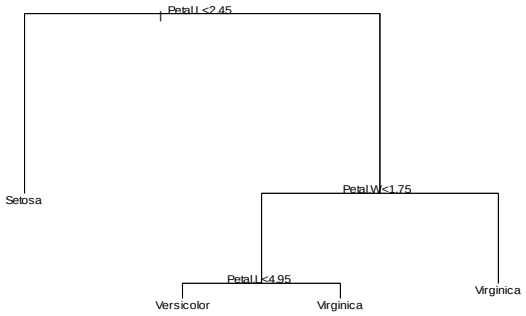
Attributes
All effects of interest
Limited WP error df
Independent β estimates



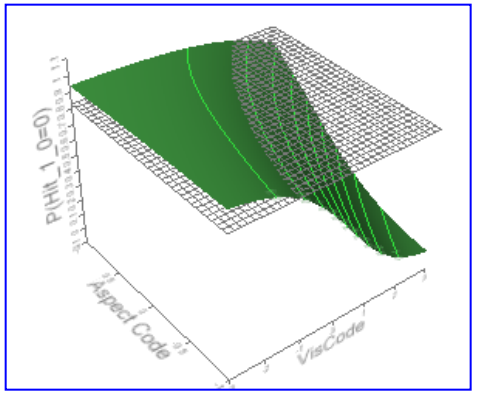
Other Methods

Modeling Alternatives

Tree-based Methods



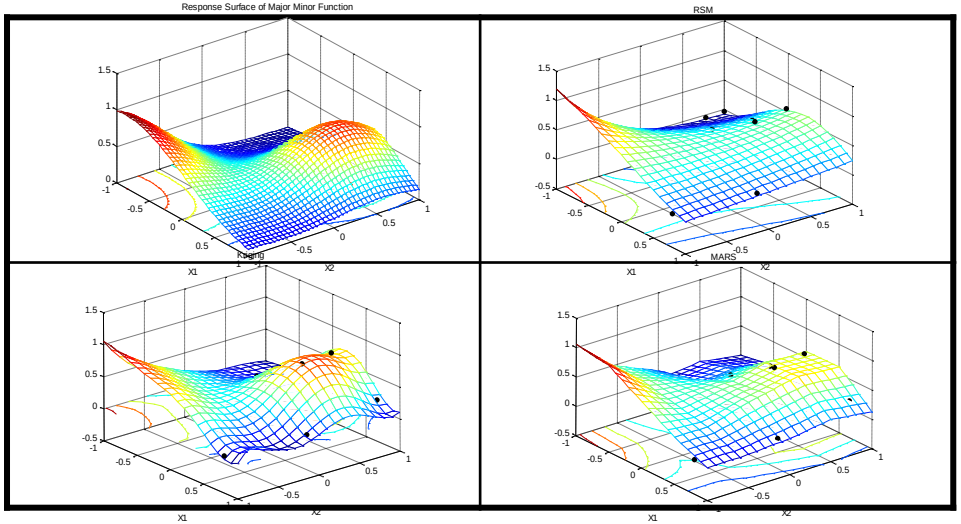
Generalized Linear Models



Nonlinear Modeling

Truth

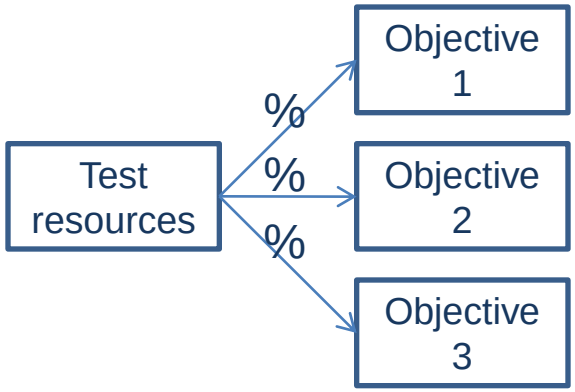
Kriging



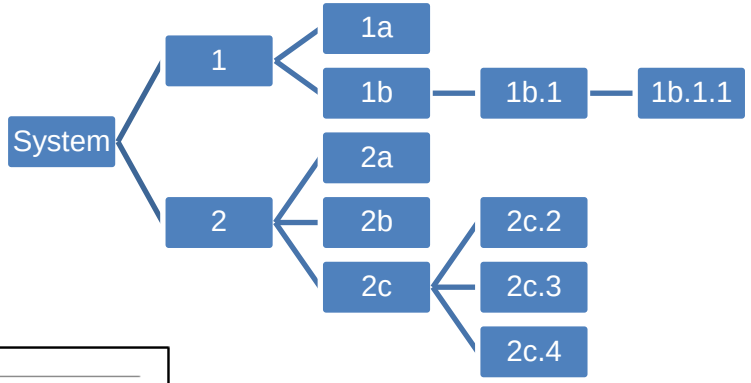
RSM

MARS

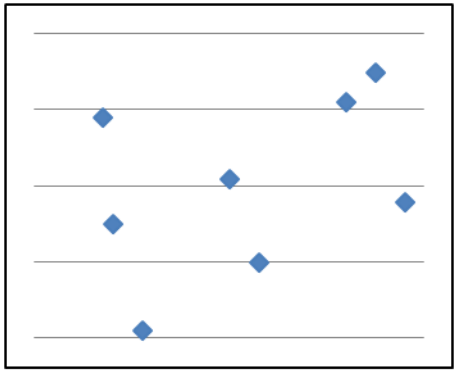
Software Testing Solutions



Decision Analysis



Factor Covering Arrays

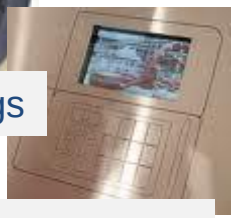


Space Filling

- **How to spread out test resources effectively/efficiently**
- **How to test configurations effectively/efficiently**
- **How to fill a space effectively/efficiently**

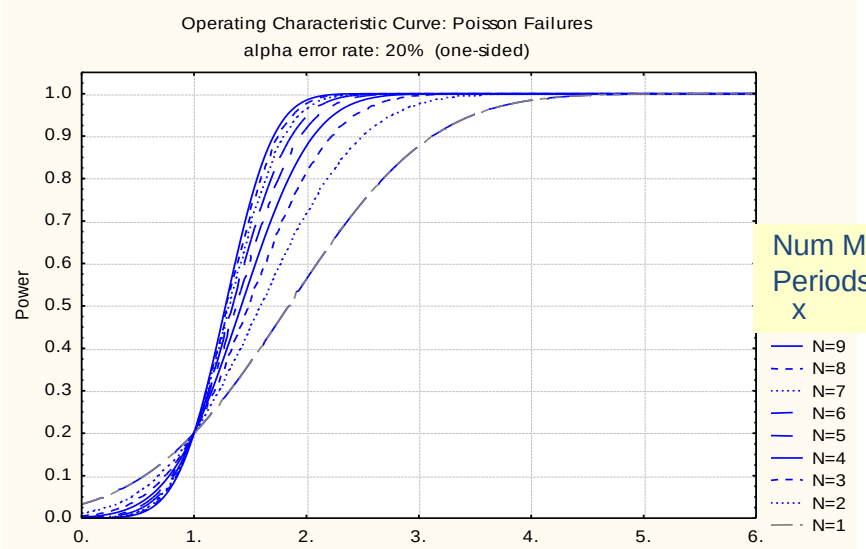
Reliability

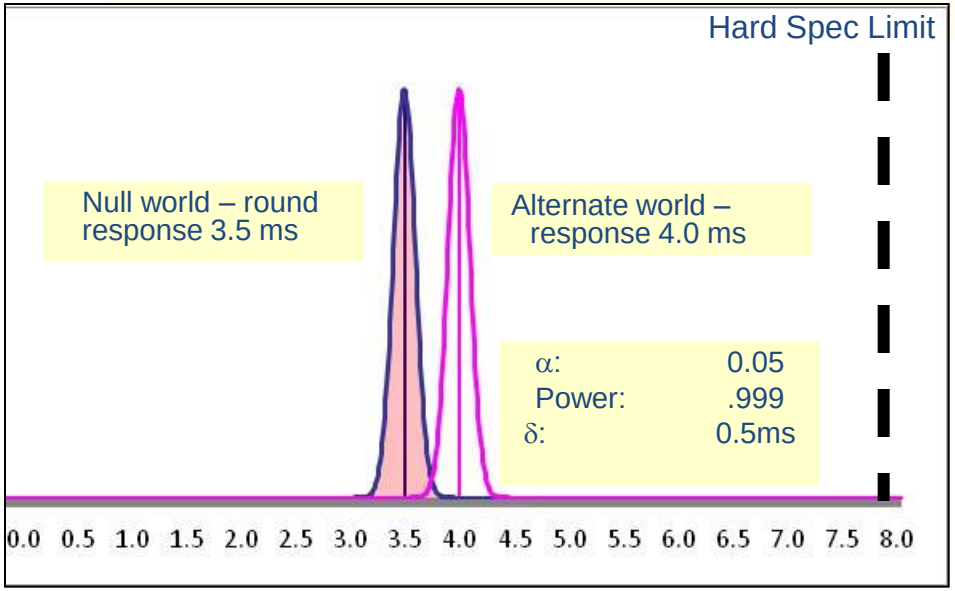
Ceramic bearings



LCD screens

GPS/INS Unit





DOE Mandate Summary

Train

- Training Program
- Mentoring – Train the Trainer
- Right Methods – Sound & Practical

Practice

- Short Term Wins – Work Projects
- Solve Tough Problems
- Research and Complement

Lead

- Leadership Commitment
- Organizational Adoption
- Metrics and Policy